

# Coherence for Frobenius pseudomonoids and the geometry of linear proofs

Lawrence Dunn  
North Florida Community College  
dunnl@nfcc.edu

Jamie Vicary  
University of Oxford  
jamie.vicary@cs.ox.ac.uk

January 20, 2016

We prove coherence theorems for Frobenius pseudomonoids and snakeoperators in monoidal bicategories. As a consequence we obtain a 3d notation for proofs in nonsymmetric multiplicative linear logic, with a geometrical notion of equivalence, and without the need for a global correctness criterion or thinning links. We argue that traditional proof nets are the 2d projections of these 3d diagrams.

## 1 Introduction

Frobenius pseudomonoids are higher-dimensional algebraic structures, first studied by Street [34], which categorify the classical algebraic notion of Frobenius algebra [24]. These higher algebraic structures have an important application to logic, since Frobenius pseudomonoids in the bicategory of categories, profunctors and natural transformations, for which the multiplication and unit have right adjoints, correspond to  $*$ -autonomous categories [4, 5], the standard categorical semantics for multiplicative linear logic. They also play a central role in topological quantum field theory [8, 9, 24, 35]. Our main result is a coherence theorem for Frobenius pseudomonoids. In the second part of the paper, we apply this coherence theorem to the problem of geometrical proof representation in linear logic, giving a 3d notation for proofs with a geometrical notion of equivalence.

### 1.1 Coherence for Frobenius pseudomonoids

In Section 2 we prove our coherence theorems. Here we give the basic definitions, and the statement of the main theorem. We begin with the definition of pseudomonoid and Frobenius pseudomonoid. These higher algebraic structures are defined in terms of *presentations*, and thanks to the graphical calculus, these presentations can be understood informally: *morphisms* are 2-dimensional tiles from which we can build composite pictures; *2-morphisms* are rewrites that allow parts of pictures to be transformed locally; and *equations* tell us that certain sequences of rewrites are equal to certain other sequences. Formally, these presentations give the generating data for finitely-presented monoidal bicategories.

**Definition 1.1.** The *pseudomonoid presentation*  $\mathcal{P}$  contains the following data:



The swallowtail equations are actually redundant, in the sense that any Frobenius pseudomonoid that does not satisfy them gives rise to one that does [3, 30], but for simplicity we include them.

We now give the statement of the coherence theorem. Let  $X$  and  $Y$  be connected, acyclic diagrams built from the generating morphisms of a Frobenius pseudoalgebra, and let  $P, Q : X \rightarrow Y$  be sequences of rewrites. Given a presentation  $\mathcal{S}$ , we write  $\mathbf{F}(\mathcal{S})$  for the free monoidal bicategory generated by this presentation [31].

**Theorem 1.3** (Coherence for Frobenius structures). *Let  $P, Q : X \rightarrow Y$  be 2-morphisms in  $\mathbf{F}(\mathcal{F})$ , such that  $X$  is connected and acyclic, with nonempty boundary. Then  $P = Q$ .*

The proof strategy is to ‘rotate’ such 2-morphisms until they are in the image of the obvious embedding  $\mathbf{F}(\mathcal{P}) \rightarrow \mathbf{F}(\mathcal{F})$ , and then to apply coherence for pseudomonoids. The central higher-categorical arguments are verified with the web-based proof assistant *Globular* [2], with direct links provided to the formalized proofs.

## 1.2 3d proofs for linear logic

In Section 3 we apply these coherence results to *multiplicative linear logic*, a formal calculus for reasoning about resources [12, 14] which is similar to traditional logic, except that resources cannot be duplicated or neglected in the way that propositions can. A central problem is determining when two proofs should be considered equivalent. In this paper, we describe a scheme for interpreting proofs as *geometrical surfaces* embedded in 3d space. We define two surfaces as *equivalent* just when one can be deformed into the other, in sense we make precise. Our main theorem then reads as follows.

**Theorem 3.2.** *Two sequent proofs in multiplicative linear logic have equal interpretations in the free  $*$ -autonomous category just when their surfaces are equivalent.*

The theory of  $*$ -autonomous categories [28, 32] is a standard mathematical model for linear logic, so this theorem says that the notion of proof equality provided by the surface calculus agrees with the standard one.

We argue that this 3d notation for linear proofs has a number of advantages over proof nets, the existing geometrical formalism for linear logic: correctness is local, with no long-trip criterion to be verified; there is no need for thinning links, and the associated non-local ‘jumps’; and it is close to categorical semantics. Furthermore, we argue that proof nets can be considered the 2d ‘shadow’ of the full 3d geometry, with the correctness criterion and thinning links arising to compensate for the fact that this ‘shadow’ has lost some essential geometrical data. We illustrate this in Figure 1, which gives a sequent proof alongside its surface diagram and proof net representations.

## 1.3 Related work

Our main theorem on coherence for Frobenius pseudomonoids can be considered a categorification of the normal form theorem for Frobenius monoids, described by Abrams [1]. Our coherence proofs are direct, and do not make use of the formalism

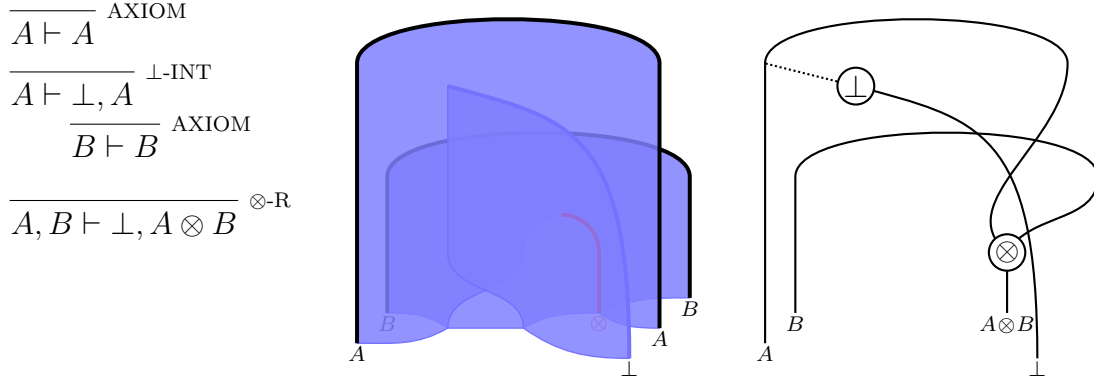


Figure 1: A deduction in the sequent calculus, and its surface calculus and proof net representations.

of rewrite systems and confluence, which has difficulties in the higher-dimensional setting [29]; when this formalism becomes better developed, recasting our proofs in such a formalism would likely be fruitful, and expose further interesting structure.

Regarding the application to logic, it was already shown by Street that Frobenius pseudomonoids with right adjoint data correspond to  $*$ -autonomous categories [34] (up to Cauchy completeness), and from work on the graphical calculus for Gray categories [6, 7, 21, 31], it is already clear that a 3d graphical calculus for  $*$ -autonomous categories must therefore exist. However, without our coherence theorem, there would be no straightforward way to understand equivalence of the 3d diagrams.

It is already well-recognized that ideas from topology are relevant to linear logic. The original proof nets of Girard [14] are topological objects, and Melliès has shown how the topology of ribbons gives a decision procedure for correctness of proof nets [27]. Proof nets allow reasoning about proofs with units, but the formalism is complex, requiring a system of thinning links with moving connections [10]. Hughes [19] gives a variant of proof nets which works well with units, but which lacks the local correctness properties of our notation. Our approach has a local flavour which is shared by the deep-inference model of proof analysis [15] and the access to monoidal coherence that it allows [18]; however, our coherence theorem is strictly more powerful, as it operates in a fragment that combines the  $\otimes$  and  $\wp$  connectives. We note also the work of Slavnov [33] on linear logic and surfaces, which involves some similar ideas to the present article, but is technically quite unrelated.

## 1.4 Acknowledgements

The authors are grateful to Samson Abramsky, Iliano Cervesato, Robin Cockett, Nick Gurski, Samuel Mimram, Robert Seely and Sam Staton for useful comments, and to the organizers of FMCS 2016, HDRA 2016 and LINEARITY 2016 for the opportunities to present these results. 3d graphics have been coded in TikZ, and 2d graphics have been produced by the proof assistant *Globular* [2].

## 2 Coherence for Frobenius pseudomonoids

### 2.1 Overview

In this section we prove the coherence theorem for Frobenius pseudoalgebras, our main technical result. The mathematical setting is higher algebraic structures in finitely-presented monoidal bicategories. We keep technicalities to a minimum, emphasizing the main ideas behind the key proof steps.

The main technical results are formalized in the web-based proof assistant *Globular* [2], and can be accessed at the address [globular.science/1601.002](http://globular.science/1601.002). To our knowledge, this is one of the first uses of *Globular* to prove a new result.

Our string diagrams in this section run bottom-to-top. Also, it is convenient to only allow *generic composites* of 1- and 2-morphisms, meaning that composite 1- and 2-morphisms have a definite length, and a linear order on their components from ‘first’ to ‘last’; this is without loss of generality (see [7] and [13, Section 2.2].) As a point of notation, we write  $\xrightarrow{\sim}$  to denote an interchanger 2-morphism, and  $\rightarrow$  to denote a composite of interchangers of length at least 0.

### 2.2 Definitions

Just as we can present a monoid by generators and relations, and a monoidal category by a PRO [26], a similar approach can be used for monoidal bicategories, as developed in the thesis of Schommer-Pries [31, Section 2.10]. Informally, a presentation can be described as follows.

**Definition 2.1.** A *presentation* of a monoidal bicategory is a list of generating objects, morphisms, 2-morphisms and equations, along with appropriate source and target data.

The sources and targets are given by composites of the data at lower levels. For example, the source and target of a generating 2-morphism is a composite of the generating objects and morphisms, under tensor product, composition and taking identities.

**Definition 2.2.** Given a presentation  $\mathcal{S}$ , we write  $\mathbf{F}(\mathcal{S})$  for the monoidal bicategory generated by this data.

In  $\mathbf{F}(\mathcal{S})$ , for 1-morphisms  $X, Y$  with the same source and target, we write  $X \equiv Y$  if they are identical as composites, and  $X \simeq Y$  if they are isomorphic using the 2-morphism structure. For 2-morphisms  $P, Q$  with the same source and target, we write  $P \equiv Q$  if they are identical as composites, and  $P = Q$  if they are equivalent modulo the equational structure.

We can now state our main theorem, which makes reference to the Frobenius presentation  $\mathcal{F}$  given in the introduction.

**Definition 2.3.** A 1-morphism in  $\mathbf{F}(\mathcal{F})$  is *connected* or *acyclic* when its string diagram graph is connected or acyclic respectively.

**Theorem 2.4** (Coherence for Frobenius structures). *Let  $P, Q : X \rightarrow Y$  be 2-morphisms in  $\mathbf{F}(\mathcal{F})$ , such that  $X$  is connected and acyclic, with nonempty boundary. Then  $P = Q$ .*

Thus, as long as one restricts to connected, acyclic diagrams with boundary, all diagrams commute. The proof strategy is to ‘rotate’ 2-morphisms in  $\mathbf{F}(\mathcal{F})$  so that they are equal to 2-morphisms in the image of the obvious inclusion  $\mathbf{F}(\mathcal{P}) \rightarrow \mathbf{F}(\mathcal{F})$ , and then rely on coherence for pseudomonoids.

### 2.3 Rotations and twistedness

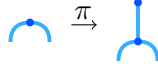
We begin by defining a new presentation, given by  $\mathcal{F}$  with some additional data added.

**Definition 2.5.** The *extended Frobenius signature*  $\mathcal{E}$  is the same as  $\mathcal{F}$ , with the following additional data:

- Morphism  $\text{cap} : \mathbf{C} \boxtimes \mathbf{C} \rightarrow \mathbf{1}$ :



- Invertible 2-morphism  $\pi$ :



- 2-Morphisms  $\sigma_1, \sigma_2, \sigma_1^{-1}$  and  $\sigma_2^{-1}$ , called the *snakeorators* and *inverse snakeorators*, equipped with additional equations as follows:

$$\begin{aligned} \sigma_1 &= \text{cap} \xrightarrow{\pi} \text{cup} \xrightarrow{\mu} \text{vertical line} & \sigma_1^{-1} &= \text{vertical line} \xrightarrow{\mu^{-1}} \text{cup} \xrightarrow{\pi^{-1}} \text{cap} \\ \sigma_2 &= \text{cup} \xrightarrow{\pi} \text{cap} \xrightarrow{\nu} \text{vertical line} & \sigma_2^{-1} &= \text{vertical line} \xrightarrow{\nu^{-1}} \text{cap} \xrightarrow{\pi^{-1}} \text{cup} \end{aligned}$$

- 2-Morphisms  $R_m, L_m, R_u, L_u, R_f, L_f$ , with additional equations as follows:

$$\begin{aligned} R_m &= \text{cup} \xrightarrow{\sigma_1^{-1}} \text{cap} \xrightarrow{\simeq} \text{cap} \xrightarrow{\pi} \text{cup} \xrightarrow{\alpha} \text{cap} \xrightarrow{\pi^{-1}} \text{cup} \\ L_m &= \text{cup} \xrightarrow{\sigma_2^{-1}} \text{cap} \xrightarrow{\simeq} \text{cap} \xrightarrow{\pi} \text{cup} \xrightarrow{\alpha^{-1}} \text{cap} \xrightarrow{\pi^{-1}} \text{cup} \\ R_u &= \text{vertical line} \xrightarrow{\sigma_1^{-1}} \text{cap} \xrightarrow{\simeq} \text{cap} \xrightarrow{\pi} \text{cup} \xrightarrow{\lambda} \text{vertical line} \\ L_u &= \text{vertical line} \xrightarrow{\sigma_2^{-1}} \text{cap} \xrightarrow{\simeq} \text{cap} \xrightarrow{\pi} \text{cup} \xrightarrow{\rho} \text{vertical line} \\ R_f &= \text{vertical line} \xrightarrow{\rho^{-1}} \text{cup} \xrightarrow{\pi^{-1}} \text{cap} \\ L_f &= \text{vertical line} \xrightarrow{\lambda^{-1}} \text{cup} \xrightarrow{\pi^{-1}} \text{cap} \end{aligned}$$

The presentations  $\mathcal{F}$  and  $\mathcal{E}$  are *equivalent*, a strong property which implies in particular that they have the same representation theory [31].

**Lemma 2.6.** *The presentations  $\mathcal{F}$  and  $\mathcal{E}$  are equivalent.*

*Proof.* This is clear, since the presentation  $\mathcal{E}$  is constructed by adding new generators to  $\mathcal{F}$ , equipped with higher cells that show how they can be expressed in terms of existing cells. In fact,  $\mathcal{E}$  is a *simple homotopy extension* of  $\mathcal{F}$ .  $\square$

The reason for constructing  $\mathcal{E}$  is that it has more convenient formal properties, which will allow us to construct our proof.

The 2-morphisms  $R_m, L_m, R_u, L_u, R_f$  and  $L_f$  can be interpreted as *rotating* their domain, either to the right or the left. For this reason we make the following definition.

**Definition 2.7.** In  $\mathbf{F}(\mathcal{E})$ , a 2-morphism is *rotational* if it is composed only from  $R_m, L_m, R_u, L_u, R_f, L_f, \sigma_1, \sigma_1^{-1}, \sigma_2, \sigma_2^{-1}$  and interchangers; we write  $\xrightarrow{R}$  to denote a composite of rotational generators of length at least 0.

**Lemma 2.8.** *In  $\mathbf{F}(\mathcal{E})$ , the snake maps satisfy the swallowtail equations:*

$$\begin{aligned} \text{id} &= \text{cup} \xrightarrow{\sigma_1^{-1}} \text{swallowtail} \simeq \text{swallowtail} \xrightarrow{\sigma_2} \text{cup} \\ \text{id} &= \text{cap} \xrightarrow{\sigma_1^{-1}} \text{swallowtail} \simeq \text{swallowtail} \xrightarrow{\sigma_2} \text{cap} \end{aligned}$$

*Proof.* Follows easily from the swallowtail equations that form part of the Frobenius presentation  $\mathcal{F}$ . See [globular.science/1601.002](http://globular.science/1601.002), 5-cells “Pf: Swallowtail 1” and “Pf: Swallowtail 2”.  $\square$

**Lemma 2.9.** *In  $\mathbf{F}(\mathcal{E})$ , the elements of each pair  $(R_m, L_m)$ ,  $(R_u, L_f)$ ,  $(L_u, R_f)$  are mutually inverse, up to interchangers and snake maps; that is, the following equations hold:*

$$\begin{aligned} \text{id} &= \text{fork} \xrightarrow{L_m} \text{snake} \xrightarrow{R_m} \text{swallowtail} \simeq \text{swallowtail} \xrightarrow{\sigma_2} \text{fork} \\ \text{id} &= \text{fork} \xrightarrow{R_m} \text{snake} \xrightarrow{L_m} \text{swallowtail} \simeq \text{swallowtail} \xrightarrow{\sigma_1} \text{fork} \\ \text{id} &= \text{cup} \xrightarrow{R_u} \text{cup} \xrightarrow{L_f} \text{snake} \simeq \text{snake} \xrightarrow{\sigma_1} \text{cup} \\ \text{id} &= \text{cup} \xrightarrow{L_u} \text{cup} \xrightarrow{R_f} \text{snake} \simeq \text{snake} \xrightarrow{\sigma_2} \text{cup} \end{aligned}$$

*Proof.* See [globular.science/1601.002](http://globular.science/1601.002), 5-cells “Pf:  $LmRm=1$ ”, “Pf:  $RmLm=1$ ”, “Pf:  $RuLf=1$ ” and “Pf:  $LuRf=1$ ”.  $\square$

**Lemma 2.10.** In  $\mathbf{F}(\mathcal{E})$ , the 2-morphisms  $\mu$  and  $\nu$  can be decomposed as follows:

$$\begin{aligned} \mu &= \text{diagram} \xrightarrow{L_m, L_f} \text{diagram} \xrightarrow{\simeq} \text{diagram} \xrightarrow{\sigma_1} \text{diagram} \xrightarrow{\lambda} \text{diagram} \\ \nu &= \text{diagram} \xrightarrow{R_m, R_f} \text{diagram} \xrightarrow{\simeq} \text{diagram} \xrightarrow{\sigma_2} \text{diagram} \xrightarrow{\rho} \text{diagram} \end{aligned}$$

*Proof.* See [globular.science/1601.002](http://globular.science/1601.002), 5-cells “Pf: *Mu Decomposition*” and “Pf: *Nu Decomposition*”.  $\square$

An important process is the elimination of adjacent cup-cap pairs in a string diagram.

**Definition 2.11.** For a 1-morphism  $X$  in  $\mathbf{F}(\mathcal{E})$ , a chosen (cup, cap) pair are *eliminable* if they match one of the following patterns:



That is, they are directly connected by a central straight wire, and turn in opposite directions. The surrounding diagram may be nontrivial, and the cup and cap may not have adjacent heights.

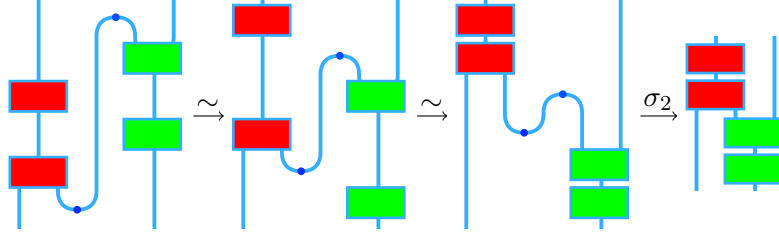
**Definition 2.12.** For a 1-morphism  $X$  in  $\mathbf{F}(\mathcal{E})$ , with a chosen eliminable cup and cap pair, the *snake removal scheme*  $X \rightarrow X'$  is the following sequence of 2-morphisms:

1. Identify the obstructing generators, which are the components of the diagram lying *between* the cup and cap in height.
2. Gather obstructing generators into maximal groups, with respect to whether they are left or the right of the central wire.
3. Interchange groups vertically beyond the cup or cap.
4. Apply the  $\sigma_1$  or  $\sigma_2$  map.

The following graphics explain the scheme. We begin with a central wire with an eliminable cup and cap. First, we identify the obstructing generators, and we gather them into groups as large as possible, which we draw in red to the left of the central wire and in green to the right, giving the first graphic below. Then we interchange



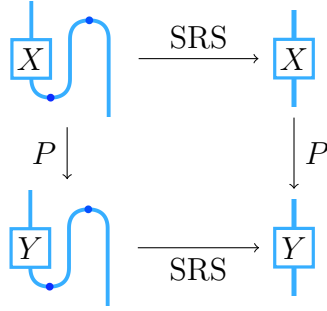
red groups up and green groups down, giving in the third graphic. Finally we cancel the cup and cap, giving the final graphic.



**Lemma 2.13.** *The snake removal scheme is natural with respect to transformations on obstructing groups.*

*Proof.* In the snake removal scheme, obstructing blocks are acted on by interchangers only, which are natural with respect to transformations of their arguments.  $\square$

**Example 2.14.** To illustrate this naturalness property, suppose that  $P : X \rightarrow Y$  is a 2-morphism in  $\mathbf{F}(\mathcal{E})$ . Then the following equation holds, where we write SRS for the 2-morphism constructed by the snake removal scheme:

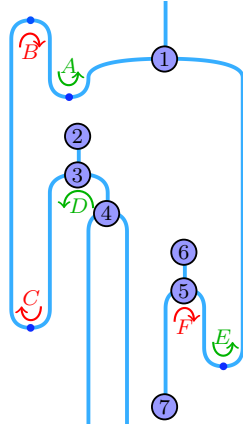


**Definition 2.15** (Connected, acyclic, simple). A 1-morphism in  $\mathbf{F}(\mathcal{E})$  is *connected* or *acyclic* when its string diagram graph is connected or acyclic, respectively; it is *simple* when it is connected and acyclic, and has a unique output wire.

**Definition 2.16** (Twistedness). Let  $X$  be a simple 1-morphism, and let  $v$  be an  $m$ ,  $u$  or  $f$  vertex in its string diagram. Then the *twistedness* of  $v$ , written  $\text{Tw}(v)$ , is the number of right turns minus the number of left turns along the shortest path from  $v$  to the unique output. We say  $X$  is *untwisted* when for all such vertices  $v$ ,  $\text{Tw}(v) = 0$ .

**Example 2.17.** Twistedness is best understood by example. The image below shows a simple 1-morphism with four  $m$  vertices (1, 3, 4, 5), one  $u$  vertex (7), and two  $f$  vertices (2, 6). The arrows show the direction of the shortest path to the output wire at the turning points, with red indicating a right turn, and green a left turn. For each of the numbered vertices, we compute the twistedness as the number

of right turns minus the number of left turns.



$$\begin{aligned}
 \text{Tw}(1) &= 0 - 0 = 0 \\
 \text{Tw}(2) &= 2 - 1 = 1 \text{ (} \textcolor{red}{C}, \textcolor{red}{B}, \textcolor{green}{A} \text{)} \\
 \text{Tw}(3) &= 2 - 1 = 1 \text{ (} \textcolor{red}{C}, \textcolor{red}{B}, \textcolor{green}{A} \text{)} \\
 \text{Tw}(4) &= 2 - 2 = 0 \text{ (} \textcolor{green}{D}, \textcolor{red}{C}, \textcolor{red}{B}, \textcolor{green}{A} \text{)} \\
 \text{Tw}(5) &= 0 - 1 = -1 \text{ (} \textcolor{green}{E} \text{)} \\
 \text{Tw}(6) &= 0 - 1 = -1 \text{ (} \textcolor{green}{E} \text{)} \\
 \text{Tw}(7) &= 1 - 1 = 0 \text{ (} \textcolor{red}{F}, \textcolor{green}{E} \text{)}
 \end{aligned}$$

Since some  $m$ ,  $u$  or  $f$  vertices have nonzero twistedness, this 1-morphism is not untwisted. Note that for computing  $\text{Tw}(2)$ ,  $\textcolor{green}{D}$  does not count as a left turn, since we are passing top-to-bottom through vertex 3; but for computing  $\text{Tw}(4)$ ,  $\textcolor{green}{D}$  does count as we are passing from bottom-right to bottom-left through vertex 3.

Untwisted simple 1-morphisms have some good properties, which we now explore.

**Lemma 2.18.** *An untwisted simple 1-morphism has no  $f$  vertices.*

*Proof.* Suppose  $v$  is an  $f$  vertex. Then travelling along the shortest path to the unique output, suppose there are  $L$  left turns and  $R$  right turns. Since the diagram is untwisted  $L = R$ , and hence  $L + R$  is an even number. Also, the only path out of an  $f$  vertex begins travelling downwards. So the path begins travelling downwards, turns left or right an even number of times, and finishes at the unique output travelling upwards; a contradiction.  $\square$

**Lemma 2.19.** *In an untwisted simple 1-morphism, every wire is either straight, or contains an eliminable (cup, cap) pair.*

*Proof.* Choose a wire  $W$ , and let  $v$  be the  $m$  or  $u$  vertex at the end closest to the unique output, and let  $v'$  be the  $m$  or  $u$  vertex at the other end. Write  $L$  for the number of left-turns along the path from  $v'$  to  $v$ , and write  $R$  for the number of right turns. Since the composite is untwisted at every vertex, we must have  $L = R$ . Suppose  $L = 0$ ; then the wire is straight. Suppose  $L > 0$ ; then there must be at least one adjacent “turn left–turn right” or “turn right–turn left” pair along the path from  $v'$  to  $v$ , and such a pair is eliminable.  $\square$

We now consider diagrams which arise from the pseudomonoid data only.

**Definition 2.20.** A 1-morphism in  $\mathbf{F}(\mathcal{E})$  is in *pseudomonoid form* when it is simple, and formed only from  $m$  and  $u$  generators; that is, just when it is connected and in the image of the natural embedding  $\mathbf{F}(\mathcal{P}) \rightarrow \mathbf{F}(\mathcal{E})$ .

**Lemma 2.21.** *A 1-morphism in pseudomonoid form is untwisted.*

*Proof.* Clearly a 1-morphism in pseudomonoid form is simple. From any  $m$  or  $u$  vertex, we build a path that travels upwards through the string diagram as far as possible. Suppose we do not reach the unique output at the top of the diagram:

then we must have encountered some component with no output legs. But that is impossible, since the 1-morphism is built solely from  $m$  and  $u$  vertices, which do have output legs. So from any vertex, there is a path to the unique output with no left or right turns; hence the diagram is untwisted.  $\square$

**Lemma 2.22.** *For every simple 1-morphism  $X$  in  $\mathbf{F}(\mathcal{E})$ , there is a rotational 2-morphism  $\Omega_X : X \xrightarrow{R} \tilde{X}$  where  $\tilde{X}$  is in pseudomonoid form.*

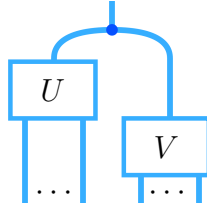
*Proof.* The generators  $R_m$ ,  $R_u$  and  $R_f$  decrease the twistedness of the diagram locally, and  $L_m$ ,  $L_u$  and  $L_f$  increase it. We apply these generators repeatedly to obtain a rotational 2-morphism  $X \rightarrow X'$ , where  $X'$  is untwisted. By Lemma 2.18,  $X'$  has no  $f$  vertices.

The 1-morphism  $X'$  is untwisted and simple, so by Lemma 2.19 every wire is either straight, or contains an eliminable cup-cap pair. We now proceed to eliminate every cup and cap, by induction on the total number of cups and caps. Suppose every wire is straight: then we are done. Otherwise, we apply the snake removal scheme of Definition 2.12 to obtain a rotational 2-morphism  $X' \rightarrow X''$ . Note that  $X''$  contains strictly fewer cups and caps than  $X'$ , and is simple since  $X'$  is, and is untwisted since it was produced from an untwisted composite by removing an adjacent left-right moving pair. So by induction, we are done, and we have eliminated all cup and cap generators.

The result is in pseudomonoid form, since by construction it is a simple 1-morphism with no  $f$ , cup or cap generators.  $\square$

We now define a stronger variant of pseudomonoid form.

**Definition 2.23.** A 1-morphism in  $\mathbf{F}(\mathcal{E})$  is in *left-pseudomonoid form* when it is identical to  $u$ , or is identical to the following composite, where  $U, V$  are in left-pseudomonoid form:



**Lemma 2.24.** *For any  $X$  in pseudomonoid form,  $\Theta_X : X \xrightarrow{\sim} \hat{X}$  where  $\hat{X}$  is in left-pseudomonoid form.*

*Proof.* We argue by induction on the number of  $m$  and  $u$  vertices. Suppose  $X \equiv u$ ; then we are done. Otherwise the uppermost vertex in  $X$  is  $m$ , and we write  $L$  and  $R$  for the sets of vertices in the subtrees below the left and right legs of  $m$  respectively. We apply interchangers to move all the vertices of  $L$  above all the vertices of  $R$ . By the inductive hypothesis, we use interchangers to arrange  $L$  and  $R$  in left-pseudomonoid form, and we are done.  $\square$

Left-pseudomonoid form is useful since it is unaltered by rotational 2-morphisms, as we now explore.

**Lemma 2.25.** *Let  $X, Y$  be in left-pseudomonoid form, with  $X \xrightarrow{R} Y$ . Then  $X \equiv Y$ .*

*Proof.* Given a simple 1-morphism, its *structure tree* is a binary tree, defined as follows. Start at the unique output, and travel along the path of the string diagram. If a  $u$  or  $f$  vertex is encountered, append a leaf to the tree. If an  $m$  vertex is encountered, append a binary branching, with subtrees given by following the other two legs of the  $m$  vertex in a similar way.

By inspection, none of the rotational 2-morphisms change the structure tree of a diagram. Also, it is clear that if two left-pseudomonoid forms have the same structure tree, they are identical. This completes the proof.  $\square$

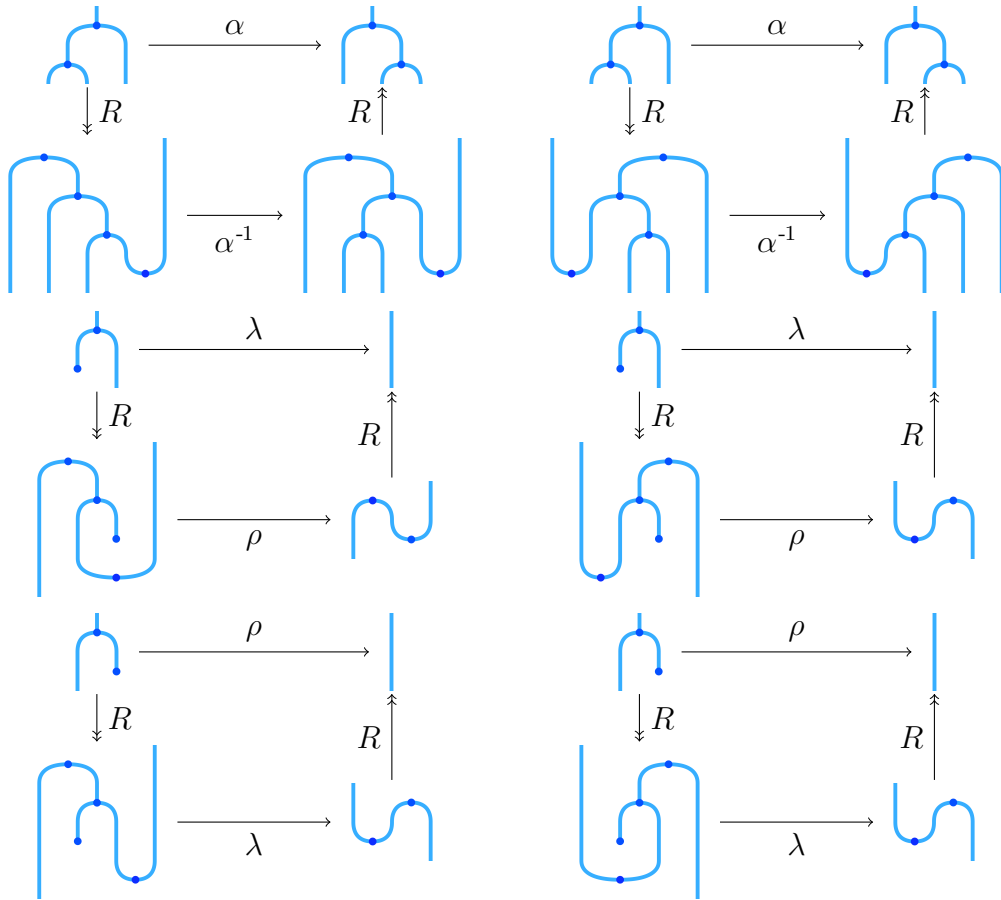
We give a name for a certain subset of the generators as follows.

**Definition 2.26.** In  $\mathbf{F}(\mathcal{E})$ , the *algebraic* generators are  $\alpha$ ,  $\lambda$ ,  $\rho$  and their inverses.

We now give the central proposition, showing how algebraic 2-morphisms can be “untwisted” to act in pseudomonoid form.

**Proposition 2.27** (Untwisting). *Let  $P : X \rightarrow Y$  be a 2-morphism in  $\mathbf{F}(\mathcal{E})$  such that  $X$  is simple. Then  $P$  is equal to a 2-morphism in which the algebraic generators act on untwisted diagrams only.*

*Proof.* We show that the following equations hold in  $\mathbf{F}(\mathcal{E})$ , where  $R$  denotes some composite of rotational generators:



See [globular.science/1601.002](http://globular.science/1601.002), “Pf: Left Alpha”, “Pf: Right Alpha”, “Pf: Right Lambda”, “Pf: Left Lambda”, “Pf: Right Rho” and “Pf: Left Rho”. These equations imply a partner set of equations where  $\alpha^{-1}$ ,  $\lambda^{-1}$  or  $\rho^{-1}$  appear at the top of each diagram.

These equations tell us that whenever an algebraic generator acts on a diagram, we can replace it by a sequence where we twist the diagram locally, apply an algebraic generator, and then twist back. A simple case analysis shows that this is sufficient to rewrite  $P$  such that any algebraic generator is acting on a locally untwisted part of the diagram. Here we analyze the possible ways that the equations above can be used to locally modify the source of an  $\alpha$ ,  $\alpha^{-1}$ ,  $\lambda$ ,  $\lambda^{-1}$ ,  $\rho$  or  $\rho^{-1}$  generator to reduce local twistedness, indicating in each picture the path to the single output wire with a dot, and giving the local twistedness as a label adjacent to each vertex:

(3)

(4)

(5)

These modifications involve the left-hand equations given above, and their partners. The twisting can be locally increased in a similar way. Thus, we can ensure that every algebraic generator acts on a part of the diagram that is locally untwisted.

However, we must show that  $P$  is equal to 2-morphism in which the *entire* diagram on which the algebraic generator acts is untwisted. Suppose  $\gamma : X \rightarrow Y$  is a single algebraic 2-morphism within the composite  $P$ , acting on a locally untwisted part of the diagram. Then by Lemma 2.13, the following equation holds:

$$\begin{array}{ccc}
 X & \xrightarrow{\gamma} & Y \\
 \Omega_X \downarrow & & \uparrow \Omega_Y^{-1} \\
 \tilde{X} & \xrightarrow{\gamma} & \tilde{Y}
 \end{array}
 \tag{6}$$

We use this to replace every algebraic generator acting on a locally untwisted part of the diagram, with an algebraic generator acting on a completely untwisted diagram.  $\square$

## 2.4 Coherence

We now give the coherence theorems.

**Proposition 2.28** (Coherence for snakeorators). *In  $\mathbf{F}(\mathcal{E})$ , let  $P, Q : X \rightarrow Y$  be composites of snakeorators, inverse snakeorators and interchangers, such that  $X$  is a connected 1-morphism. Then  $P = Q$ .*

*Proof.* We define  $S : Y \rightarrow Y'$  by applying the snake-removal scheme of Definition 2.12 to each wire of  $Y$ , and we define  $P' = S \circ P$  and  $Q' = S \circ Q$ . Clearly  $P' = Q'$  just when  $P = Q$ . Next, note that  $Q'$  is invertible; we define  $P'' = P' \circ Q'^{-1} : Y' \rightarrow Y'$ , and clearly  $P'' = \text{id}_{Y'}$  just when  $P' = Q'$ . So without loss

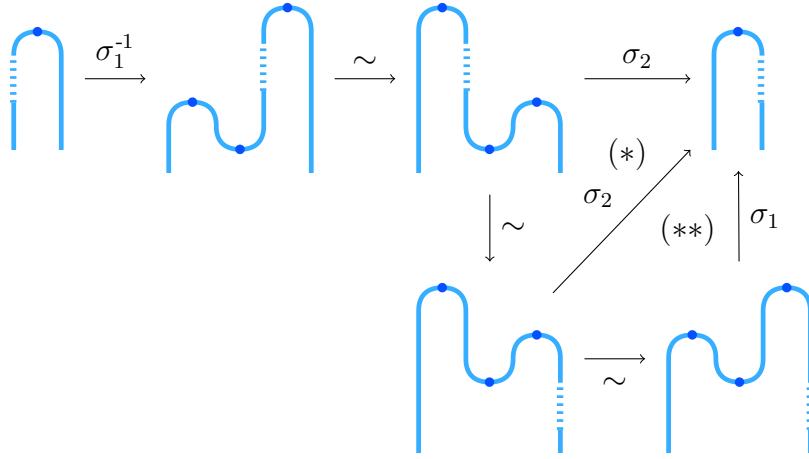


Figure 2: Turning  $\sigma_2$  into  $\sigma_1$  in a general composite

of generality, we may assume that  $X \equiv Y$  and  $Q = \text{id}$ , and that  $X$  has no eliminable cup-cap pairs in the sense of Definition 2.11; our task is then to show that  $P = \text{id}$ .

The composite  $P$  is formed from  $\sigma_1$ ,  $\sigma_2$ ,  $\sigma_1^{-1}$ ,  $\sigma_2^{-1}$  and interchanger maps. The source of the  $\sigma_1^{-1}$  and  $\sigma_2^{-1}$  maps is the identity, so by naturality, they can all be moved to the beginning of the composite. Similarly, the  $\sigma_1$  and  $\sigma_2$  maps can all be moved to the end of the composite. So the rearranged composite is of the following form: a sequence of inverse snakeorators, followed by a sequence of interchangers, followed by a sequence of snakeorators.

We argue by induction on the number of inverse snakeorator generators. This proof will have two inductive arguments: we call this first one  $I_1$ . Suppose there are no inverse snakeorators: then we are done. Otherwise, write  $F$  for the final inverse snakeorator present in the rearranged composite. Suppose for the rest of the proof that it is a  $\sigma_1^{-1}$  generator; for a  $\sigma_2^{-1}$  generator, a similar proof applies.

The final inverse snakeorator  $F$  produces a single cup and cap, at least one of which will be annihilated later in the composite, since by assumption  $X$  has no eliminable cup-cap pairs. There are two ways that this can happen: either (A) they are annihilated with each other by a  $\sigma_1$  generator, or (B) one of them is annihilated with a neighbouring cup or cap by a  $\sigma_2$  generator. Suppose case (B) holds; then we will show our composite is equal to one in which (A) in fact holds. The argument is illustrated in Figure 2. We begin with some composite that contains a cap. At some point lower down the left-hand wire, a  $\sigma_1^{-1}$  generator acts to create a new cup and cap pair. Later in the composite, following the application of interchangers, and snakeorators to other cups and caps not displayed, a  $\sigma_2$  generator is applied to the right-hand cup and cap. (We are supposing that it is the newly-created *cup* which is annihilated first; if the newly-created cap is annihilated first, a reflected argument applies.) In triangle (\*), we observe by naturality that the composite is equal to one in which the  $\sigma_2$  is applied higher up, adjacent to the leftmost cap. In triangle (\*\*), we apply one of the swallowtail equations to replace the  $\sigma_2^{-1}$  with a  $\sigma_1^{-1}$  that annihilates the same cup and cap created in the first arrow of the diagram.

We assume therefore that case (A) holds. Local to the cup and cap produced by

$F$ , the diagram will look like this:

$$\left| \xrightarrow{\sigma_1^{-1}} \text{cup} \xrightarrow{\sim, \sigma_1, \sigma_2} \text{cap} \xrightarrow{\sigma_1} \right| \quad (7)$$

The central arrow represents some composite of interchangers,  $\sigma_1$  and  $\sigma_2$  generators. By assumption, the  $\sigma_1$  and  $\sigma_2$  maps will act only on the rest of the diagram, not the displayed cups and caps, but the interchangers may affect any part of the diagram. Since this central arrow is not the identity, we cannot immediately apply the invertibility equation to cancel the  $\sigma_1^{-1}, \sigma_1$  pair from the composite.

During the course of this central arrow, the chosen cup and cap may become separated vertically, with the cap always remaining above the cup.<sup>1</sup> Since the initial and final separation is clearly 0, this separation must have some maximum value  $n \in \mathbb{N}$  over the course of the central arrow. We now show that our composite is equal to one for which this maximum value is 0. We give an inductive argument  $I_2$ , jointly on the value  $n$  of this maximum, and its multiplicity  $m$  over the course of the central arrow (since the separation may reach its maximum value of  $n$  more than once over the composite.)

There are 4 ways that the separation between the cup and cap can increase, by applying interchangers as follows, where the boxes  $A$  and  $B$  indicate arbitrary composites, the box  $C$  is a composite of height 1 (i.e. a generator padded on the left and right), and the black wires are arbitrary types:

$$\begin{array}{c} \text{C} \\ | \\ \text{A} \end{array} \text{---} \text{---} \text{B} \rightarrow \begin{array}{c} \text{C} \\ | \\ \text{A} \end{array} \text{---} \text{---} \text{B} \quad (8)$$

$$\begin{array}{c} | \\ \text{A} \end{array} \text{---} \text{---} \text{B} \begin{array}{c} \text{C} \\ | \end{array} \rightarrow \begin{array}{c} | \\ \text{A} \end{array} \text{---} \text{---} \text{B} \begin{array}{c} \text{C} \\ | \end{array} \quad (9)$$

$$\begin{array}{c} | \\ \text{A} \end{array} \text{---} \text{---} \text{B} \begin{array}{c} \text{C} \\ | \end{array} \rightarrow \begin{array}{c} | \\ \text{A} \end{array} \text{---} \text{---} \text{B} \begin{array}{c} \text{C} \\ | \end{array} \quad (10)$$

$$\begin{array}{c} | \\ \text{A} \end{array} \text{---} \text{---} \text{B} \begin{array}{c} \text{C} \\ | \end{array} \rightarrow \begin{array}{c} | \\ \text{A} \end{array} \text{---} \text{---} \text{B} \begin{array}{c} \text{C} \\ | \end{array} \quad (11)$$

<sup>1</sup>The cap could only go below the cup if an additional inverse snakeoperator  $\sigma_2^{-1}$  was applied to the central connecting wire, but by assumption  $\sigma$  is the final inverse snakeoperator present in the composite.

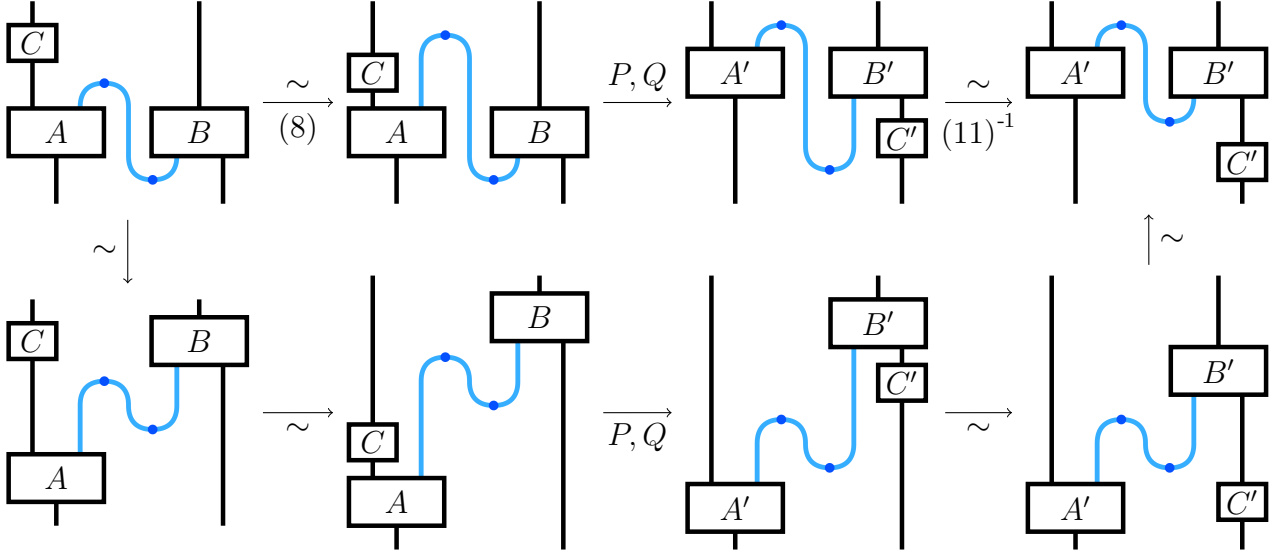


Figure 3: Reducing the local maximum separation between a cup and cap

In each case the separation increases by 1 unit. After this increase, arbitrary interchangers can take place in the diagram, not involving the cup and cap; this will leave the separation invariant. Finally, the separation will decrease, giving the end of the local maximum. It could decrease by the reverse of one of the processes (8–11) (4 ways); or by a snakeorator being applied in a region between the cup and cap in height, on the left or the right of the central blue wire (2 ways).

So in total there are 4 ways for the separation to increase, and 6 ways for it to decrease, giving a total of 24 types of local maximum. For each case, suppose the separation increases from  $n$  to  $n + 1$ , and then decreases from  $n + 1$  to at most  $n$ . Then we can show directly that this is equal to a composite in which the separation is at most  $n$  throughout. While the number of cases is large, they are all handled straightforwardly and in a similar way.

We analyze one case in detail. Suppose that the separation increases from  $n$  to  $n + 1$  by method (8); followed by 2-morphisms  $P, Q$  applied between the cup and cap in height, to the left and right of the central wire respectively; followed by the separation decreasing from  $n + 1$  to at most  $n$  by method (11). (2-morphisms applied above the cap or below the cup may be moved later in the composite by naturality, and we neglect them.) Then by the argument in Figure 3, we can perform the operations in a different order, under which the separation is at most  $n$ . Commutativity of this diagram, and others like it for the different cases, follows from naturality of the interchanger in a monoidal bicategory.

By this argument, we have reduced the multiplicity of local maxima of the separation by 1; or, if the multiplicity was already 1, we have reduced the global maximum. By the hypothesis of the inductive argument  $I_2$ , we are done.

We have demonstrated that the composite in (7) is equal to one in which we apply  $\sigma_1^{-1}$ , act only on the cups and caps by interchanging them as a single unit, and then apply  $\sigma_1$ . By naturality of the interchanger, the  $\sigma_1$  and  $\sigma_1^{-1}$  can now be cancelled. Thus we have shown that our composite is equal to one with strictly fewer inverse snakeorators. By the hypothesis of the inductive argument  $I_1$ , we are done.  $\square$



**Proposition 2.29** (Coherence for rotational 2-morphisms). *In  $\mathbf{F}(\mathcal{E})$ , let  $X$  be a simple 1-morphism, and let  $P, Q : X \xrightarrow{R} Y$  be rotational 2-morphisms. Then  $P = Q$ .*

*Proof.* The rotational 2-morphisms are invertible, so  $P = Q$  just when  $Q^{-1} \circ P = \text{id}_X$ ; so it is enough to consider the case that  $Y = X$  and  $Q = \text{id}_X$ . Also, by Lemma 2.22, there is a rotational isomorphism  $\Omega_X : X \rightarrow \tilde{X}$ , where  $\tilde{X}$  is in pseudomonoid form; so without loss of generality, we may suppose that  $X$  is in pseudomonoid form, and therefore untwisted by Lemma 2.21.

Let  $v$  be any  $m$  vertex in  $X$ . Since  $R_m$  and  $L_m$  maps act locally on multiplication vertices, and no rotational cells can introduce or eliminate  $m$  vertices, we can move all  $R_m$  and  $L_m$  instances acting on  $v$  to the beginning of the composite  $P$ . Since the source and target of  $P$  is untwisted, an equal number of  $R_m$  and  $L_m$  2-morphisms must act on it, since they change the twistedness locally; we cancel adjacent  $(R_m, L_m)$  and  $(L_m, R_m)$  pairs using Lemma 2.9, at the cost of introducing additional snakeorators. Similarly, each  $u$  vertex in  $X$  will be rotated by a succession of  $R_u$ ,  $L_u$ ,  $R_f$  and  $L_f$  maps; as for the  $m$  vertices, we move these to the beginning of the composite, and eliminate them pairwise. Hence we obtain  $P = P' : X \rightarrow X$ , where  $P'$  is formed purely of interchangers and snake maps. But then by Proposition 2.28, we conclude  $P' = \text{id}$ .  $\square$

**Lemma 2.30.** *In  $\mathbf{F}(\mathcal{E})$ , suppose  $X, Y$  are 1-morphisms in pseudomonoid form, and  $X \xrightarrow{R} Y$ . Then  $X \xrightarrow{\sim} Y$ .*

*Proof.* Consider the following diagram:

$$\begin{array}{ccc} X & \xrightarrow{R} & Y \\ \Theta_X \searrow & & \nearrow \Theta_Y^{-1} \\ & \hat{X} = \hat{Y} & \end{array} \quad (12)$$

The 2-morphisms  $\Theta_X$  and  $\Theta_Y$  are defined by Lemma 2.24, and are composed purely of interchangers. The equality at the bottom follows from Lemma 2.25. By Proposition 2.29, the diagram commutes. Since the lower path  $X \xrightarrow{\sim} \hat{X} = \hat{Y} \xrightarrow{\sim} Y$  is composed purely of interchangers, the result follows.  $\square$

We now prove our main result.

**Theorem 2.4.** (Coherence for Frobenius structures.) *Let  $P, Q : X \rightarrow Y$  be 2-morphisms in  $\mathbf{F}(\mathcal{F})$ , such that  $X$  is connected and acyclic, with at least one boundary wire. Then  $P = Q$ .*

*Proof.* Let  $X$  be a simple 1-morphism in  $\mathbf{F}(\mathcal{F})$ , and let  $P, Q : X \rightarrow Y$  be 2-morphisms. Since the 2-morphisms of  $\mathbf{F}(\mathcal{F})$  are invertible, then  $P = Q$  just when  $Q^{-1} \circ P = \text{id}$ ; so it is enough to consider the case that  $X \equiv Y$  and  $Q = \text{id}$ . We then proceed as follows.

- Define  $P_1 : X \rightarrow X$  as the image of  $P$  under the embedding  $\mathbf{F}(\mathcal{F}) \rightarrow \mathbf{F}(\mathcal{E})$ . It is enough for us to prove  $P_1 = \text{id}$ , since by Lemma 2.6, the presentations  $\mathcal{F}$  and  $\mathcal{E}$  are equivalent.
- Define  $P_2$  by taking  $P_1$  and eliminating all instances of  $\mu$  and  $\nu$  using Lemma 2.10.

- Define  $P_3$  by taking  $P_2$  and ensuring all algebraic generators act on untwisted diagrams only, using Proposition 2.27.
- Define  $P_4 := \Omega_X \circ P_3 \circ \Omega_X^{-1} : \tilde{X} \rightarrow \tilde{X}$ .

Clearly  $P_1 = P_2 = P_3$ , and  $P_3 = \text{id}$  just when  $P_4 = \text{id}$ .

We now consider the structure of  $P_4$ . Its source and target are in pseudomonoid form, and  $P_4$  itself is built from rotational 2-morphisms, along with algebraic 2-morphisms acting on diagrams of pseudomonoid form. That is,  $P_4$  is of the following form, where  $X_i$  and  $Y_i$  are of pseudomonoid form,  $\gamma_i$  are algebraic 2-morphisms, and  $R$  indicates a composite rotational 2-morphism:

$$P_4 = X_0 \xrightarrow{R} Y_0 \xrightarrow{\gamma_0} X_1 \xrightarrow{R} Y_1 \xrightarrow{\gamma_1} \cdots \xrightarrow{R} Y_n = X_0$$

By applying Lemma 2.30, we can build  $P_5$  as follows, with  $P_4 = P_5$ :

$$P_5 = X_0 \xrightarrow{\sim} Y_0 \xrightarrow{\gamma_0} X_1 \xrightarrow{\sim} Y_1 \xrightarrow{\gamma_1} \cdots \xrightarrow{\sim} Y_n = X_0$$

This composite lies purely in the image of the pseudomonoid presentation  $\mathcal{P}$ . But then by the coherence theorem for pseudomonoids, we conclude  $P_5 = \text{id}$ ; see the work of Lack [25], and the thesis of Houston [17, Section 6].

In the statement of the theorem, we required only that  $X$  had a single input and output. To reduce to the case of a unique output, one can compose with cups and caps appropriately; the swallowtail equations ensure this gives a bijection of hom-sets.  $\square$

## 2.5 Adjoints

We now consider the case that the generators  $m, u$  have adjoints.

**Definition 2.31.** The *right-adjoint Frobenius presentation*  $\mathcal{F}^*$  is defined to be the Frobenius presentation  $\mathcal{F}$ , with the following additional data:

- Additional 1-morphisms  $m^*$  and  $u^*$ :

$$\text{Diagram of } m^* \text{ (a cup with a red dot)} \quad \text{Diagram of } u^* \text{ (a cap with a red dot)} \quad (13)$$

- Additional 2-morphisms  $\eta, \epsilon, \phi, \psi$ :

$$\text{Diagram of } \eta \text{ (two vertical lines to a cup)} \quad \text{Diagram of } \epsilon \text{ (a cup to two vertical lines)} \quad (14)$$

$$\text{Diagram of } \phi \text{ (a vertical line to a red dot)} \quad \text{Diagram of } \psi \text{ (a red dot to a vertical line)} \quad (15)$$

- Additional equations, stating that  $m \dashv m^*$  and  $u \dashv u^*$ :

$$\text{id} = \text{Diagram of } m \text{ followed by } m^* \text{ (cup and cap)} \quad \text{Diagram of } \eta \text{ followed by } \epsilon \text{ (two vertical lines)} \quad (16)$$

$$\text{id} = \text{Diagram of } u \text{ followed by } u^* \text{ (cap and cup)} \quad \text{Diagram of } \phi \text{ followed by } \psi \text{ (vertical line)} \quad (17)$$

**Definition 2.32.** In  $\mathbf{F}(\mathcal{F}^*)$ , define the following composites:

$$\begin{array}{c} \text{blue wire with red dot} \end{array} := \begin{array}{c} \text{blue wire with red dot} \end{array} \quad \begin{array}{c} \text{red dot} \end{array} := \begin{array}{c} \text{blue wire with red dot} \end{array} \quad (18)$$

**Lemma 2.33.** In  $\mathbf{F}(\mathcal{F}^*)$ , we have the following adjunctions:

$$\begin{array}{cc} \text{blue wire with red dot} & \dashv & \text{blue wire with red dot} \\ \text{blue wire with red dot} & \dashv & \text{blue wire with red dot} \end{array}$$

*Proof.* The first row of adjunctions are explicit in the definition of  $\mathcal{F}^*$ . The second row are constructed straightforwardly from the available data.  $\square$

## 2.6 Graphical calculus for Frobenius pseudomonoids

As a result of the coherence theorem, we can relax our conventions for the graphical calculus for a Frobenius pseudomonoid. For any number of input and output wires (as long as there is at least one in total), we may draw a generic vertex that connects them:



By the coherence theorem, we are guaranteed that all acyclic, connected ways of forming a composite with this type will be canonically isomorphic, so we may as well draw this simpler vertex to represent the entire isomorphism class. Furthermore, given any two connected acyclic 1-morphisms with boundary in  $\mathbf{F}(\mathcal{F})$ , there is a canonical choice of morphism between them, which in the surface calculus we can show as a pointlike operation, such as expression (19).

For the right-adjoint Frobenius presentation  $\mathcal{F}^*$ , we have additional graphical components available thanks to the adjunctions described in Lemma 2.33.

## 2.7 \*-Autonomous categories

From the work of Street [34], it is known that the following definition of non-symmetric  $*$ -autonomous category agrees with that of Barr [5], up to a Cauchy-completeness assumption. Here **Prof** is the symmetric monoidal bicategory of categories, profunctors and natural transformations [11].

**Definition 2.34.** A  $*$ -autonomous category is a pseudomonoid  $(\mathbf{C}, m, u, \alpha, \lambda, \rho)$  in **Prof**, equipped with a morphism  $f : \mathbf{C} \rightarrow \mathbf{1}$  such that  $f \circ m : \mathbf{C} \times \mathbf{C} \rightarrow \mathbf{1}$  is a biexact pairing, and such that  $m$  and  $u$  have right adjoints.

**Proposition 2.35.** The data of a  $*$ -autonomous category is equivalent to that of an  $\mathcal{F}^*$ -structure in **Prof**.

*Proof.* The reverse direction immediate, since given an  $\mathcal{F}^*$ -structure in **Prof**, the composite  $f \circ m$  is clearly biexact, thanks to the invertible 2-morphisms  $\mu$  and  $\nu$ .

For the forward direction, we sketch the main idea. In Definition 2.34, the biexact pairing condition means that if one composes the right leg of  $f \circ m : \mathbf{C} \times \mathbf{C} \rightarrow \mathbf{1}$

with the Hom-profunctor of type  $\mathbf{1} \rightarrow \mathbf{C} \times \mathbf{C}^{\text{op}}$ , the resulting profunctor of type  $\mathbf{C} \rightarrow \mathbf{C}^{\text{op}}$  is an equivalence. Therefore, this definition is in terms of *two* objects,  $\mathbf{C}$  and  $\mathbf{C}^{\text{op}}$ , and structures defined on them. But since  $\mathbf{C}$  and  $\mathbf{C}^{\text{op}}$  are equivalent, it stands to reason that we can transport any structures defined on  $\mathbf{C}^{\text{op}}$  across to  $\mathbf{C}$ , yielding a definition in terms of  $\mathbf{C}$  alone. This is an algorithmic procedure, and the result is an  $\mathcal{F}^*$ -algebra in **Prof**.  $\square$

## 3 3d proofs for linear logic

### 3.1 Overview

In this section we develop a theory of 3d surface proofs for multiplicative linear logic without the exchange rule. Our surfaces are similar in spirit to *proof nets* [10, 14], graphical structures which are traditionally used extensively to denote linear proofs. However, we argue that our scheme has several advantages:

1. Correctness is local; any well-typed composite produces a valid proof-theoretic object, with no global property (such as the *long-trip criterion* [14]) to be verified.
2. Equivalence is local, unlike for proof nets where the main dynamical rewiring step involves non-local jumps, and requires the re-validation of a global property [10, Section 3.1].
3. Equivalence is broad, establishing some proof equivalences in fewer steps than for proof nets; sometimes in just one step.
4. It works well with the unital fragment of the logic, which is problematic for traditional proof nets, requiring decorations in the form of *thinning links* [10].
5. It is compositional, in the sense that the surface for a composite proof is just the union of the surfaces for any partition.
6. It is close to categorical semantics, with a surface giving rise straightforwardly to a morphism in a  $*$ -autonomous category.

Nonetheless, we propose that proof nets are essentially 2d *projections* of the 3d surface geometry. From this perspective, we can make sense of some of the features of proof nets: the long-trip criterion can be interpreted as a non-local check that the 2d shadow is consistent with a valid 3d geometry, and the thinning link decorations indicate the depth at which a unit is attached in the 3d geometry. We illustrate this in Figure 1, which gives a sequent calculus deduction along with its surface calculus and proof net representations.

We criticise our scheme in the following ways.

1. When we say in point 4 above that it works well with the unital fragment of the logic, we mean that the notion of proof equivalence is straightforward and local, not that one obtains an efficient method for deciding proof equality [16].
2. We do not offer a decision procedure for proof equality in terms of our 3d notation.

3. We work in a non-symmetric variant of the logic which lacks the exchange rule; adding a symmetry would mean that the surfaces self-intersect, which would complicate the formalism, although we expect our results could be extended.
4. As complex geometrical objects, our surfaces are not as easy to manipulate with pen-and-paper as traditional proof nets; to become practical, advances in proof assistant technology (such as [2]) would be required.

### 3.2 The 2d calculus

In this section we develop the 2d string diagram calculus for sequents. This calculus is the Joyal-Street calculus for monoidal categories [22], directed from left to right. We use the standard 2-sided sequent calculus for nonsymmetric multiplicative linear logic with units [10]: our sequents are pairs  $\Gamma \vdash \Delta$ , where  $\Gamma$  and  $\Delta$  are ordered lists (separated with “,”) of expressions in the following grammar, where  $V = \{A, B, C, \dots\}$  is a set of atomic variables:

$$S ::= \{I, \perp\} \mid V \mid S \otimes S \mid S \wp S \mid S^* \mid *S$$

We have left and right negation, we assume isomorphisms  $*(S^*) \simeq S \simeq (*S)^*$  which, for simplicity, we generally suppress at the syntactic level. Atomic variables are represented as black dots, pointing in different directions depending on their side of the sequent:

$$\begin{array}{cc} A \bullet \cdots & \cdots \bullet A \\ A \vdash \cdots & \cdots \vdash A \end{array}$$

The two sides of a sequent are represented graphically by trees, which are drawn connected together at their roots. The basic connective “,” is denoted as a blue vertex with zero or more branches to the left or right, as follows:

$$\begin{array}{ccc} \begin{array}{c} A \bullet \\ B \bullet \end{array} \text{---} \begin{array}{c} \bullet C \\ \bullet D \\ \bullet E \end{array} & \begin{array}{c} \bullet \text{---} A \\ \bullet \text{---} B \end{array} & \begin{array}{c} A \bullet \\ B \bullet \\ C \bullet \end{array} \text{---} \bullet \\ A, B \vdash C, D, E & \vdash A, B & A, B, C \vdash \end{array}$$

The connectives  $\otimes$  and  $\wp$ , which are always binary, are drawn in blue on their natural side (left for  $\otimes$ , right for  $\wp$ ), and in red on the other side, as we show with the following examples:

$$\begin{array}{ccc} \begin{array}{c} A \bullet \\ B \bullet \\ C \bullet \\ D \bullet \end{array} \text{---} \begin{array}{c} \bullet E \\ \bullet F \\ \bullet G \\ \bullet H \end{array} & \begin{array}{c} A \bullet \\ B \bullet \\ C \bullet \end{array} \text{---} \begin{array}{c} \bullet D \\ \bullet E \\ \bullet F \end{array} & \\ A, (B \otimes C), D \vdash E, (F \otimes G), H & (A \wp B), C \vdash D \wp (E \otimes F) & \end{array}$$

Note that a blue dot with a binary branching is therefore an overloaded notation; this is a deliberate feature.

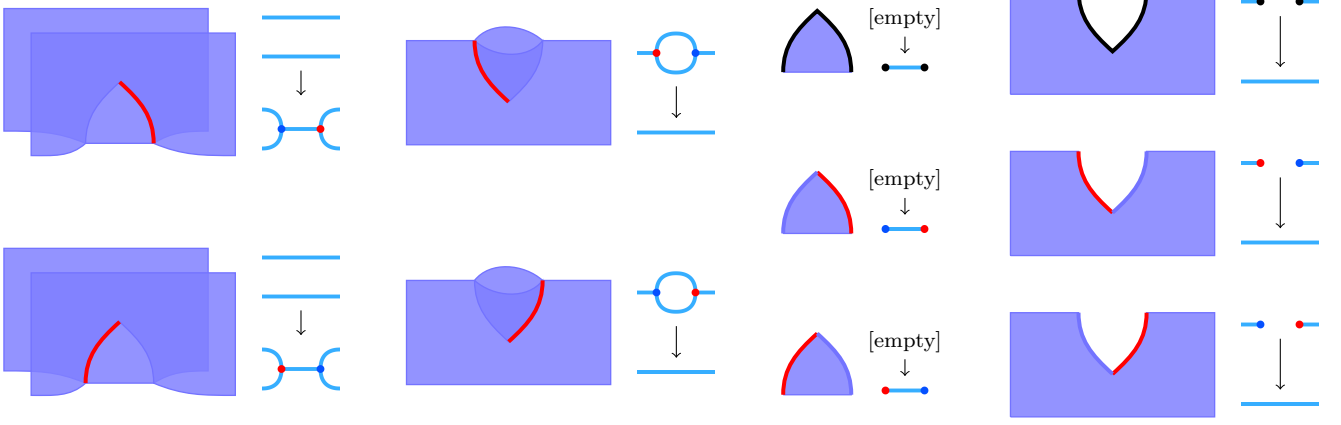


Figure 4: The generators of the 3d calculus

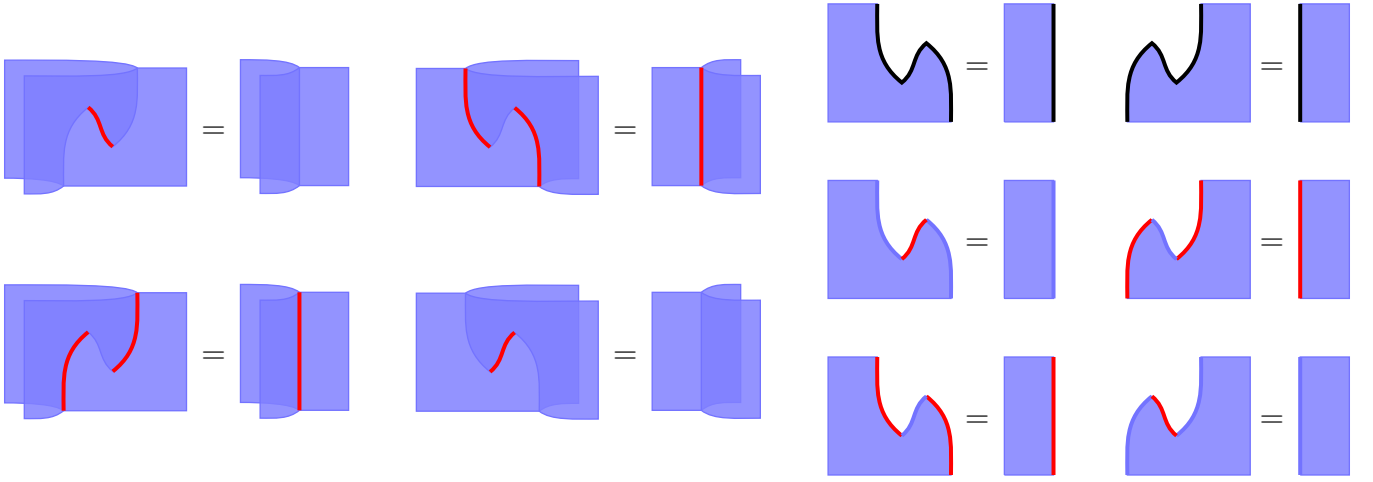


Figure 5: The equations of the 3d calculus

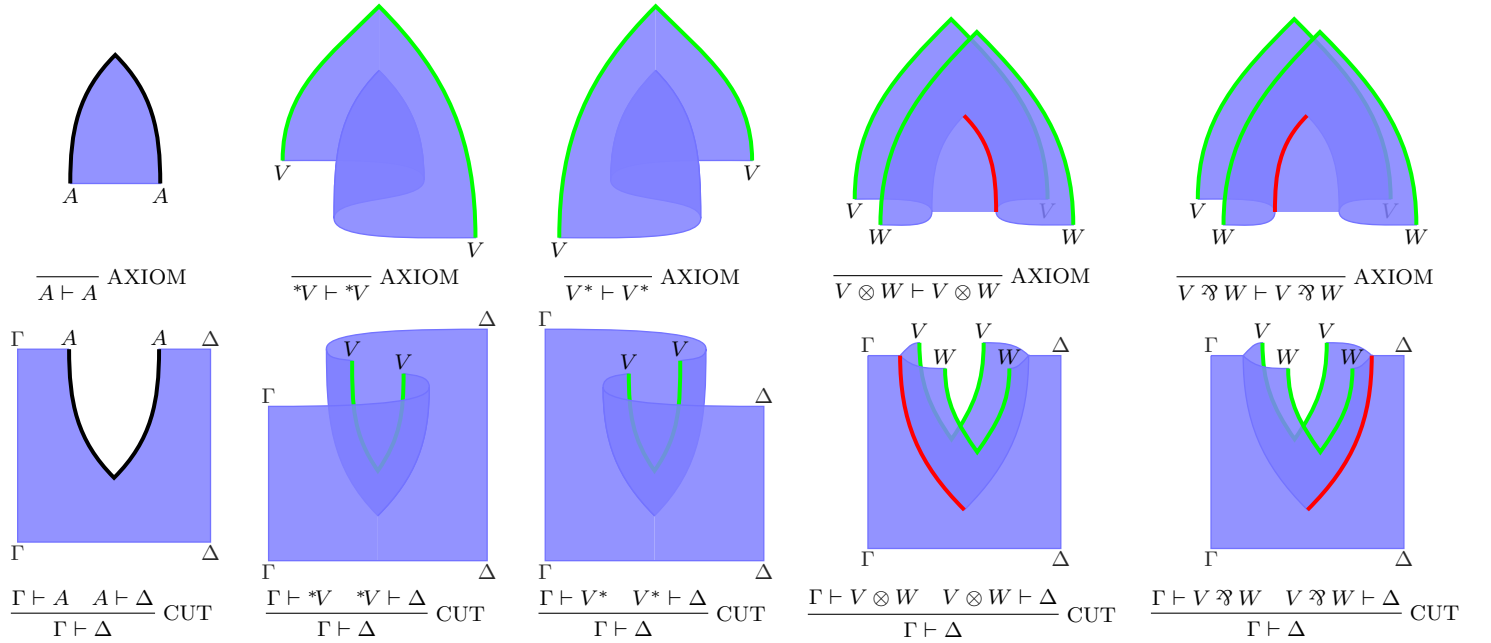
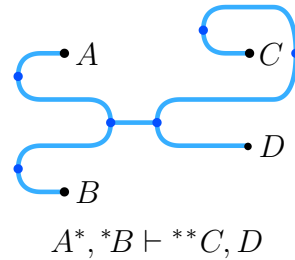


Figure 6: AXIOM and CUT rules for an atomic variable  $A$ , and recursively for variables  $V, W$

The units  $I$  and  $\perp$  are represented by blue dots on their natural side (left for  $I$ , right for  $\perp$ ), and red dots on the other side, as shown:



We represent  $(-)^*$  as turning right by a half-turn, and  $*(-)$  as turning left by a half-turn, as shown:

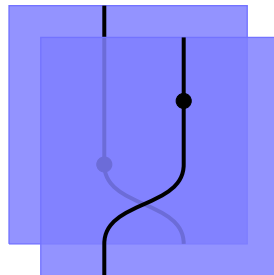


Diagrams built from sequents in this way are of a simple kind; as graphs, they are all acyclic and connected. In general we can allow arbitrary well-typed composites of these components; such diagrams represent 1-morphisms in the monoidal bicategory  $\mathbf{F}(\mathcal{F}^*)$ , described in Section 2.

### 3.3 The 3d calculus

Diagrams in the 3d calculus are surfaces embedded in  $\mathbb{R}^3$ . Formally they are expressions in the graphical calculus for Gray categories, which is by now well-developed [6, 7, 21, 31]. However, the 3d calculus is quite intuitive, and we take advantage of this to introduce it in an informal way.

Diagrams consist of *sheets*, bounded on the left and right by *edges*, which are bounded above and below by *vertices*. (Sheets can also be bounded by the sides of the diagram, and edges can also be bounded by the top or bottom of the diagram.) Diagrams are immersed in 3d space, meaning that sheets can exist in front or behind other sheets, and wires on sheets of different depths can cross; however, components never intersect. Here is an example:



Here we have front and back sheets, each containing an edge, which contains a vertex. Towards the bottom of the picture, the wires cross: this is called an *interchanger*.

For our application to linear logic, we allow two types of vertex: *coherent vertices* and *adjunction vertices*.

- **Coherent vertices.** Say that a 2d calculus diagram is *simple* when it is connected and acyclic with nonempty boundary, and in the blue fragment of the calculus, not involving red nodes or black atomic variable nodes. Then any two simple diagrams can be connected by a coherent vertex, denoted as follows:



On the left we give the surface representation, and on the right we give the 2d calculus representation of the upper and lower boundaries. The coherent vertex is the point in the middle of the surface diagram where 4 edges meet.

- **Adjunction vertices.** Listed in Figure 4, these introduce and eliminate red and black edges in the surface calculus.

We now define equivalence in the graphical language, giving intuitive interpretations of each generating relation in *italics*.

**Definition 3.1.** Two surface diagrams are *equivalent* when they are related by the least equivalence relation generated by the following:

- **Coherence.** Let  $P, Q$  be surface diagrams built from interchangers and coherent vertices, whose upper and lower boundaries are simple 2d diagrams; then  $P = Q$ . (*All acyclic equations of coherent vertices hold.*)
- **Adjunction.** The equations listed in Figure 5 hold. (*Bent wires can be pulled straight.*)
- **Isotopy.** The equations of a monoidal bicategory hold. (*If two diagrams are ambient isotopic, they are equivalent.*)
- **Locality.** Suppose surface diagrams  $P, Q$  differ only with respect to subdiagrams  $P', Q'$ , with  $P' = Q'$ . Then  $P = Q$ . (*Equivalence applies locally in the interior of a diagram.*)

It is a fair summary of this definition to say that two diagrams are equivalent just when one can be *deformed* into the other.

Our presentation here is informal, but we emphasize that our definition of the surface calculus and its equivalence relation can be made completely precise in terms of the formal development of Section 2: two diagrams are equivalent just when they are equal as 2-morphisms in the monoidal bicategory  $\mathbf{F}(\mathcal{F}^*)$ .



### 3.4 Interpreting the sequent calculus

We saw in Section 3.2 how individual sequents in multiplicative linear logic can be interpreted as 2d diagrams. We now see how proofs can be interpreted as 3d surface diagrams. We view these surfaces as directed from top to bottom, just like traditional sequent calculus proofs; so for a particular surface, its *hypothesis* is the upper boundary, and its *conclusion* is the lower boundary.

We use a basis for the sequent calculus with a symmetry between introduction and elimination for  $\otimes$ ,  $\wp$ ,  $I$  and  $\perp$ ; the rules  $\otimes$ -R,  $\wp$ -L,  $I$ -R and  $\perp$ -L are derivable (see Example 3.4.) The interpretation of AXIOM and CUT rules are given recursively in Figure 6; the interpretation of the remaining rules, which we call the *virtual fragment* of the logic, is given in Figure 7.

We now establish the main theorem that controls proof equivalence for this 3d notation. We now prove our main theorem regarding the application to linear logic.

**Theorem 3.2.** *Two sequent proofs in multiplicative linear logic have equal interpretations in the free  $*$ -autonomous category just when their surfaces diagrams are equivalent.*

*Proof.* The generators in Figure 4, and the equations in Figure 5, are precisely the definition of an  $\mathcal{F}^*$ -structure rendered in the surface calculus. (In fact they are redundant, since they include additional adjunction equations which Lemma 2.33 shows are already derivable.) Under the equivalence of Proposition 2.35, it is essentially immediate that this reduces to the standard interpretation of the sequent calculus in a  $*$ -autonomous category [32]: CUT is by composition and AXIOM is the identity morphism, mediated by the adjunctions  $V_* \dashv V^*$  of profunctors induced by a variable defined by a functor  $V : \mathbf{1} \rightarrow \mathbf{C}$ ;  $(\wp, \perp)$  is the monoidal product of the  $*$ -autonomous category; the negations come from the biclosed structure; and  $(\otimes, I)$  is the dual monoidal structure induced by negation.  $\square$

It is interesting to analyze the different contributions to proof equivalence made by each part of Definition 3.1 of surface equivalence. **Coherence** tells us that any two proofs built in the virtual part of the logic given in Figure 7 are equal. **Adjunction** tells us that AXIOM and CUT cancel each other out, both for atomic and compound variables. **Isotopy** tells us that that ‘commutative conversion’ is possible, where we exchange the order of multiple independent hypotheses. **Locality** tells us that we can apply our equations in the context of a larger proof, in the manner of deep inference [15].

We give a formal statement of coherence for the virtual fragment of the logic, since it is a result of independent interest.

**Corollary 3.3.** *If two sequent proofs in the virtual fragment of the logic given in Figure 7 have the same hypotheses and conclusion, then they are equal in the free  $*$ -autonomous category.*

We understand this to be a folklore result, in the sense that it seems to be known to some members of the community, but we have been unable to identify a proof in the literature. At best, our corollary therefore fills a gap in the literature; at least, we offer a new method of proof for this result.

We comment on some interesting features of the translation between the sequent calculus and the surface calculus. The fundamental simplicity of the surface calculus

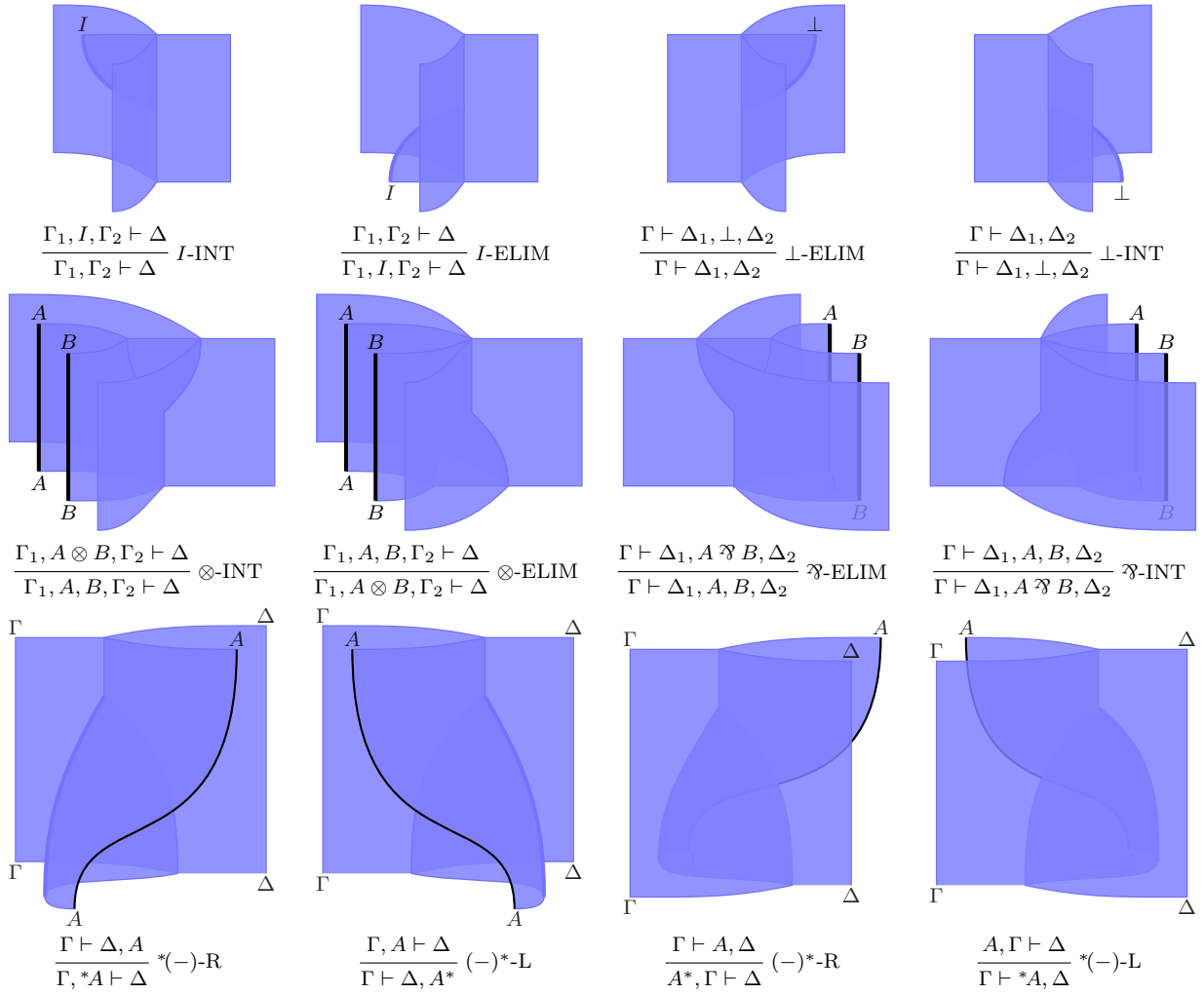


Figure 7: Surface interpretations of the virtual part of the logic

is clear, from the minimality of the data in Figure 4, as compared to Figures 6 and 7. Partly this is achieved by the greater degree of locality: for example, the cut rules for  $*V$  and  $V^*$  are both interpreted using the same surface generators, composed in different ways. But more significantly, the entire virtual fragment of the sequent calculus is interpreted in the *trivial* part of the surface calculus, significantly reducing the bureaucracy of proof analysis, to use Girard's phrasing [14]. To make the most of these advantages, we suggest that the surface calculus can serve directly as a toolkit for logic, not just as a way to visualize sequent calculus proofs.

### 3.5 Examples

We now look in detail at a number of examples: we derive the surface form of the missing  $\otimes\text{-R}$  rule; we analyze equivalence of a proof involving units; and we investigate the classic triple-dual problem.

**Example 3.4** (Additional rules). Presentations of multiplicative linear logic usually include the rules  $\otimes\text{-R}$ ,  $\wp\text{-L}$ ,  $I\text{-R}$ ,  $\perp\text{-L}$ , which are missing from Figure 6 and Figure 7; however, they are derivable. We analyze  $\otimes\text{-R}$  in detail in Figure 8. On the left-hand side, we derive the rule in our chosen basis for the sequent calculus. In the middle

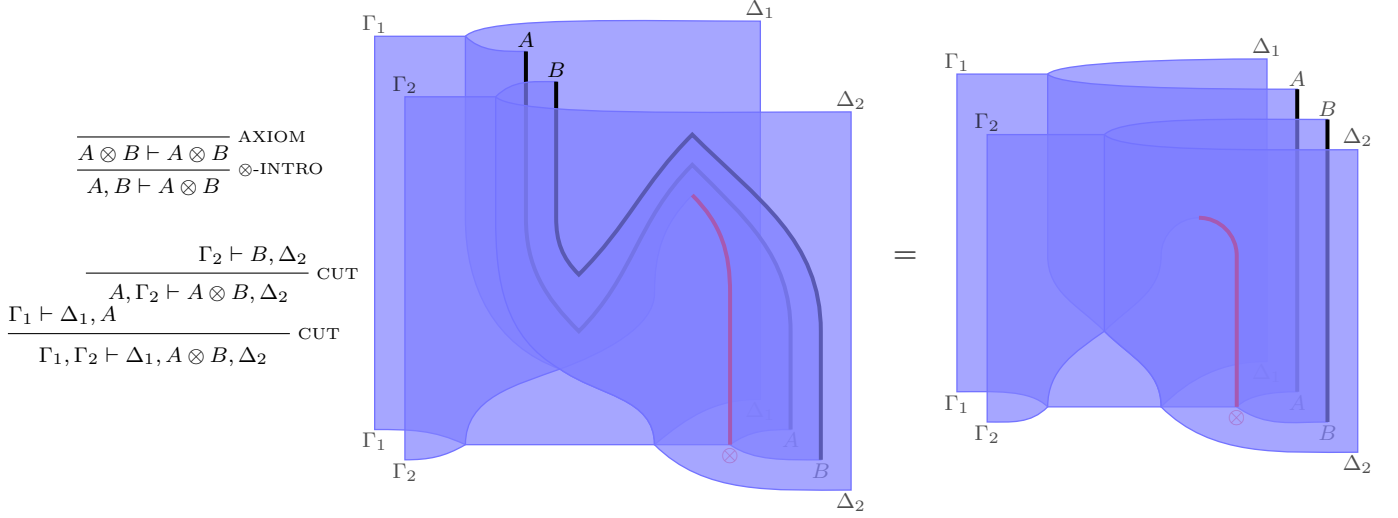


Figure 8: Derivation of the surface representation of the  $\otimes$ -R rule

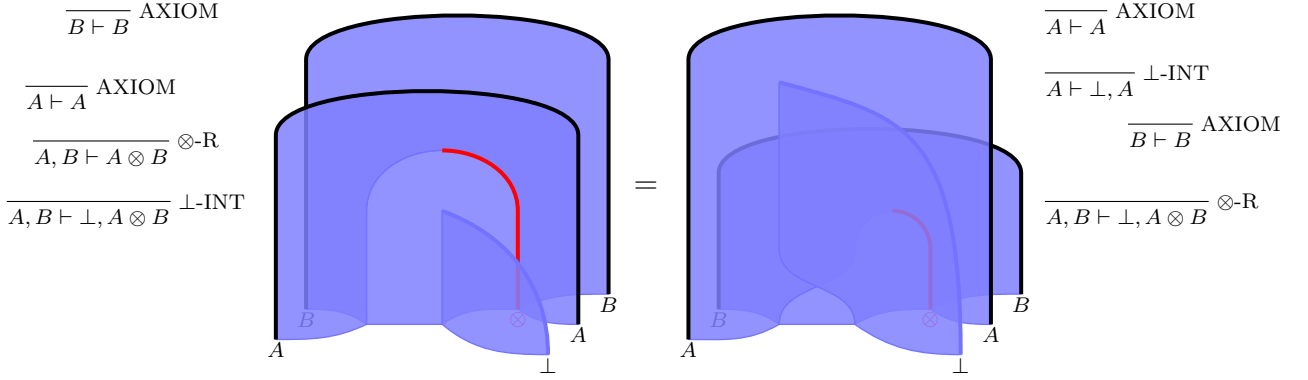


Figure 9: Geometrical equivalence of two proofs involving the unit

image, we interpret it in the surface calculus, using the rules we have described. In the third image, we simplify the surface calculus interpretation using the rules in Figure 5. From this simplified diagram, we see that it does not in fact involve the variables, the nontrivial generators being applied in the central part of diagram only. Elegant interpretations of the other 3 rules can be derived similarly.

**Example 3.5** (Equivalence of two proofs with units). The example is given in Figure 9. We present two distinct sequent proofs of the tautology  $A, B \vdash \perp \mathcal{A}(A \otimes B)$ , along with their corresponding surface proofs. The heights are aligned to help understand how the surface proofs have been constructed. We make use of the  $\otimes$ -R rule derived in the previous example.

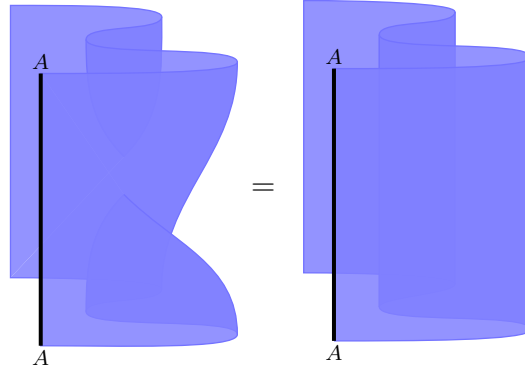
It can be seen by inspection that the surface proofs are equivalent, as follows. Starting with the surface on the left, we allow the  $\perp$ -introduction vertex to move up and to the left; this is an application of **Coherence** and **Locality**. We also allow the  $B$ -introduction vertex at the top of the diagram to move down, behind both the  $A$ -introduction and  $\perp$ -introduction vertices; this is an application of **Isotopy**.

**Example 3.6** (Triple-dual problem). Starting with the identity  $A \multimap X \rightarrow A \multimap X$ , we can uncurry on the left to obtain a morphism  $A \otimes (A \multimap X) \rightarrow X$ , and curry on

the right to obtain a morphism  $p_A : A \rightarrow X \multimap (A \multimap X)$ ; in a similar way, we can also define a morphism  $q_A : A \rightarrow (X \multimap A) \multimap X$ . Then the *triple-dual problem*, originally due to Kelly and Mac Lane [23] and generalized here to the non-symmetric setting, is to determine whether the following equation holds:

$$\begin{array}{ccc}
 X \multimap ((X \multimap A) \multimap X) & \xrightarrow{X \multimap q_A} & X \multimap A \\
 & \searrow \text{id} & \downarrow p_{X \multimap A} \\
 & & X \multimap ((X \multimap A) \multimap X)
 \end{array} \tag{20}$$

Choosing  $X = \perp$ , then  $q_A$  and  $p_A$  are the isomorphisms  $*(A^*) \simeq A \simeq (*A)^*$ . We give the surfaces for the clockwise and anticlockwise paths of (20):



$$\tag{21}$$

We conclude that the proofs are equivalent by a *single* application of the **Coherence** rule. This problem is a standard test case, and we can contrast our one-step deduction with the treatment of Blute et al [10, Section 4.2] and Hughes [20, Example 2] in terms of proof nets, where the proofs require several rewiring steps.

## References

- [1] L. Abrams. Two-dimensional topological quantum field theories and Frobenius algebras. *Journal of Knot Theory Ramifications*, 05(05):569–587, oct 1996.
- [2] K. Bar, A. Kissinger, and J. Vicary. Globular: an online proof assistant for higher-dimensional rewriting. [ncatlab.org/nlab/show/Globular](http://ncatlab.org/nlab/show/Globular), 2016.
- [3] K. Bar and J. Vicary. Promoting adjoint 1-morphisms in a Gray category to satisfy the swallowtail equations. Formal proof at [globular.science/1512.006](http://globular.science/1512.006).
- [4] M. Barr. \*-autonomous categories and linear logic. *Math. Struct. in Comp. Science*, 1(02):159, jul 1991.
- [5] M. Barr. Nonsymmetric \*-autonomous categories. *Theoretical Computer Science*, 139(1-2):115–130, mar 1995.
- [6] J. W. Barrett, C. Meusburger, and G. Schaumann. Gray categories with duals and their diagrams. *J. Diff. Geom.*, to appear. arXiv:1211.0529.

- [7] B. Bartlett. Quasistrict symmetric monoidal 2-categories via wire diagrams. 2014. arXiv:1409.2148.
- [8] B. Bartlett, C. L. Douglas, C. Schommer-Pries, and J. Vicary. Extended 3-dimensional bordism as the theory of modular objects. 2014. arXiv:1411.0945.
- [9] B. Bartlett, C. L. Douglas, C. Schommer-Pries, and J. Vicary. Modular categories as representations of the 3-dimensional bordism 2-category. 2015. arXiv:1509.06811.
- [10] R. Blute, J. Cockett, R. Seely, and T. Trimble. Natural deduction and coherence for weakly distributive categories. *JPAA*, 113(3):229–296, 1996.
- [11] F. Borceux. *Handbook of Categorical Algebra: Volume 1, Basic Category Theory*. Cambridge University Press, 1994.
- [12] R. Di Cosmo and D. Miller. Linear logic. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2015 edition, 2015.
- [13] C. Douglas and A. Henriques. Internal bicategories. arXiv:1206.4284, 2012.
- [14] J.-Y. Girard. Linear logic. *TCS*, 50(1):1–101, 1987.
- [15] A. Guglielmi. *All about Proofs, Proofs for All*, chapter Deep Inference, pages 164–172. College Publications, 2015. Online.
- [16] W. Heijltjes and R. Houston. No proof nets for MLL with units. In *Proceedings of CSL-LICS 2014*.
- [17] R. Houston. *Linear logic without units*. PhD thesis, University of Manchester, 2007. arXiv:1305.2231.
- [18] D. Hughes. Deep inference proof theory equals categorical proof theory minus coherence. Unpublished note. Online, 2004.
- [19] D. Hughes. Simple multiplicative proof nets with units. arXiv:math/0507003, 2005.
- [20] D. Hughes. Simple free star-autonomous categories and full coherence. *JPAA*, 216(11):2386–2410, 2012.
- [21] B. Hummon. *Surface diagrams for Gray categories*. PhD thesis, UC San Diego, 2012. Online.
- [22] A. Joyal and R. Street. The geometry of tensor calculus, I. *Adv. Math.*, 88(1):55–112, 1991.
- [23] G. M. Kelly and S. MacLane. Coherence in closed categories. In *Saunders Mac Lane Selected Papers*, pages 471–514. 1979.
- [24] J. Kock. *Frobenius Algebras and 2D Topological Quantum Field Theories*. Cambridge University Press (CUP), 2003.
- [25] S. Lack. A coherent approach to pseudomonads. *Adv. Math.*, 152(2):179–202, 2000.

- [26] S. Lack. Composing PROPs. *TAC*, 13(9):147–163, 2004. Online.
- [27] P.-A. Melliès. A topological correctness criterion for multiplicative non-commutative logic. In *Linear Logic in Computer Science*, pages 283–322. Cambridge University Press (CUP).
- [28] P.-A. Melliès. *Interactive models of computation and program behaviour*, chapter Categorical Semantics of Linear Logic. 2009.
- [29] S. Mimram. Towards 3-dimensional rewriting theory. *Logical Methods in Computer Science*, 10(2), apr 2014.
- [30] P. Pstragowski. On dualizable objects in monoidal bicategories. Master’s thesis, Bonn University, 2014. arXiv:1411.6691.
- [31] C. Schommer-Pries. *The Classification of Two-Dimensional Extended Topological Field Theories*. PhD thesis, Department of Mathematics, University of California, Berkeley, 2009. arXiv:1112.1000.
- [32] R. A. G. Seely. *Categories in Computer Science and Logic*, chapter Linear logic, \*-autonomous categories and cofree coalgebras, pages 371–382. AMS, 1989.
- [33] S. Slavnov. From proof-nets to bordisms: the geometric meaning of multiplicative connectives. *MSCS*, 15(06):1151, 2005.
- [34] R. Street. Frobenius monads and pseudomonoids. *J. Math. Phys.*, 45(10):3930, 2004.
- [35] U. Tillmann. *S*-Structures for  $k$ -linear categories and the definition of a modular functor. *Journal of the London Mathematical Society*, 58:208–228, 1998.